AUSTRALIAN HATE CRIME NETWORK

Submission to the Australian Government's Consultations on a new Online Safety Act

21 February 2020

Submitted on behalf of the Australian Hate Crime Network by:
Associate Professor Nicole L Asquith
Secretary, Australian Hate Crime Networ
N.Asquith@westernsydney.edu.au
02 4736 0951

INTRODUCTION

We thank the Australian Government for this opportunity to contribute to a vital area of law reform. As well as responding to questions posed in the Government's discussion paper, this submission outlines modest proposals for how Australia's online safety framework could evolve to deal with the immense private and public harm caused by generalised public advocacy of hatred and violence against segments of the community.

This content of this submission represents the views of non-government and academic members of the Australian Hate Crime Network. It does not represent the views of representatives of any government agency or department.

EXECUTIVE SUMMARY

There are many positive aspects to the proposed reforms, and we commend the Australian Government's leadership in this area. From a social policy perspective however, there are significant ways in which these reforms can be strengthened.

In 2019, the e-Safety Commissioner's office undertook research in partnership with Europe and New Zealand to understand the impacts of online hate speech.[1] From the 3700+ Australians surveyed, the overwhelming majority supported action to check the spread of online hate speech, including the introduction of legislation and getting social media companies to do more.

Victoria University research has identified an 'increasingly radical milieu' that is leaning closer to violent outcomes. A Victorian analysis of far right extremist activity online[2], including over 41,000 posts in 12 far right Facebook pages, identified dangerous narratives that continue to thrive on digital platforms without detection. For example, the narrative that Muslims are 'inferior, sub-human, and inherently incompatible with Western liberal norms and values'; or the narrative that 'gender fluidity and same sex marriage would open the gates for normalising paedophilia, polygamy, beastiality [sic] and incest.' Material that advances these extremist narratives are dangerous and need to be consistently removed from digital platforms.

While there is a porous relationship between online hatred and extremist movements, policy parameters need to recognise that approaching this from a *countering violent extremism* angle alone will not address hate crime. This is largely attributable to the fact that the majority of public acts of hatred both offline and online are committed by ordinary people without membership or affiliation to a far right group.

Public advocacy of hatred and violence has very real consequences, and erodes security, safety and wellbeing for targeted groups both online and offline. The AHCN has proposed a greater articulation of the **higher objectives** of the Act to include recognition of both private and public harms; the pivotal importance of protecting community confidence in online spaces as safe places; and upholding the parity of protections from crime, whether it happens online or offline.

The AHCN supports the **regulatory statement of policy** with a couple of specific options for amendment that correspond with the suggested higher objectives.

Instead of the proposed **Basic Online Safety Expectations (BOSE)**, the AHCN recommends the Government consider developing a **Continuum of Online Safety Expectations (COSE)** in consultation with industry and key community stakeholder representatives (the COSE is explained further below). Continuums are a superior policy tool to minimum expectations, in

---

1 Australian Government, e-Safety Commissioner, *Online Hate Speech: Findings from Australia, New Zealand and Europe* (January 2020), <https://www.esafety.gov.au/about-us/research/online-hate-speech>.

2 Dr Mario Puecker, Dr Debra Smith, & Dr Muhammad Iqbal, 'Mapping Networks and Narratives of Far-Right Movements in Victoria' (Project Report, Institute for Sustainable Industries and Liveable Cities, Victoria University, November 2018).

fields which aim to foster and incentivise constant innovation and improvement. In developing the continuum, expectations and consequences align with the location of an entity on the continuum, particularly in terms of the size of its user base. High profit margins correspond with greater capacity and would also elevate the level of expectation and responsibility for continual improvement.

In regard to the **ambit of the Online Safety Charter and COSE/BOSE** (the key policy instruments put forward in these reforms):

- They should both incentivise social media service providers and online services to **enforce an adequate standard in relation to rejecting public advocacy of hatred or violence.** The Online Safety Act should incorporate a new legislative standard for public advocacy of hatred and violence that applies to all groups, which we know from empirical evidence, are targeted by hate crime. This legal standard should also be acknowledged within the Online Safety Charter of community expectations; and embedded and enforced through the BOSE/COSE as a minimum expectation. This submission seeks to generate much-needed discussion about the need for such legislation by including possible wording for such a provision.
- Greater **stakeholder policy engagement** in continual improvement of platform self-regulation and moderation processes must be encouraged in the COSE/BOSE. For example, see the post Christchurch government-industry Taskforce's recommendation to involve academia and civil society in highlighting evolving depictions of extremist content.
- The COSE/BOSE should recognise that **reliance on platform based appeal mechanisms** in the context of public advocacy of hatred or violence is inappropriate and harmful to users, noting the onerous burden this places on victims of target communities.
- The COSE/BOSE also need to incentivise continual improvement in relation to **handling of repeat offenders**, achieving a standard that maintains community confidence in our justice system and the safety of online spaces.
- The COSE/BOSE need to cover online service/platform's **cooperation with removing barriers to justice**, including specific tools for quickly determining if an offender is in the same jurisdiction; and identifying the location of repeat offenders.
- The COSE/BOSE should develop strategies to reduce **false news stories** about segments of the community that are used to perpetuate hatred; remedying/correcting the information when identified; and enacting consequences for responsible account holders.
- The COSE/BOSE should aim to improve **the operation of algorithms** and other processes that may drive users towards (or amplify) terrorist and extreme violent material to better understand possible intervention points, and to implement changes where this occurs.
- BOSE/COSE should streamline expectations for industry into one document, appropriately recognising the porous relationship between online hatred and extremism.

In regard to appropriate **sanctions,** this Submission argues that it is reasonable and necessary to penalise platforms for failing to comply with the BOSE/COSE. The AHCN submits it would be reasonable to align expectations and consequences within a continuum that reflects provider capacity. This means that social media service providers at the top of the continuum, who have the practical and financial capabilities to achieve a high level of safety but fail to do so, would

face penalties. Platforms that continually allow criminal breaches or refuse to implement strategies for moderation in line with the Online Safety Charter, thereby choosing to operate at the lowest end of the online safety continuum (COSE), should face the risk of being deplatformed. Meanwhile, implementation of significant improvements in line with BOSE/COSE could reduce financial penalties.

The AHCN supports the concept of **transparency reports** but submits that public advocacy of hatred or violence must be included within the scope of 'illegal, abusive or predatory' content, amongst other recommendations.

The AHCN supports the alignment of and proposed improvements to the **cyber bullying (children) and cyber abuse (adults) schemes,** with the addition of in-built processes in the COSE/BOSE for: acknowledging providers with the fastest take-down rates; ensuring that incidents that have a dimension of prejudice-motivated hatred are identified and treated as an aggravating factor, not only for data integrity but also restorative justice and victim healing; and for clarifying the victim's recourse where the Commissioner refuses to take down material.

The AHCN submits that **'illegal and harmful content' should include the public advocacy of hatred or hate-related violence, and that take-down powers be conferred for extreme examples that are clearly harmful.** It proposes that consideration be given to two alternative routes to implement this: either the definition of cyber-abuse be extended to cover abuse expressed to whole groups of people on the basis of their identity; or the civil standard proposed in this document be used as the threshold at which the Commissioner can issue a take-down notice. Where there are multiple requests for take down notices against the same platform, this should constitute grounds for further sanctions as per the COSE/BOSE.

In relation to the **blocking of terrorist or extremely violent content** online, the e-Safety Commissioner needs to ensure that extremist manifestos connected to terror attacks causing loss of life, are RC classified. The AHCN asks the Government to consider what appeal mechanisms should be available if the e-Safety Commissioner refuses to exercise this power. In addition, the AHCN seeks further information and engagement on the proposed development of a tool kit for civil society to respond to extremist ideologies online. To the extent civil society is conferred this role, it is also noted resourcing will be required.

Finally, the AHCN recommends that the **Commissioner's role** be clarified to include facilitating the removal of barriers to justice online; as well as conducting research into the effectiveness of international rehabilitation programs for repeat offenders of online hatred, for consideration by Australian jurisdictions. Reducing recidivism is closely aligned with prevention.

REFORM OBJECTIVES

The AHCN support the objectives of this reform process to[3]:

1. Maintain the elements of the existing framework that are working well, such as the cyberbullying and image-based abuse schemes
2. Address gaps in current regulatory arrangements, particularly where the current schemes are out of date or don't address harms occurring on more recently developed services or platforms.
3. Establish a more flexible framework that can accommodate new online harms as they emerge.
4. Hold the perpetrators of harmful online conduct accountable for their actions online
5. Improve the transparency and accountability of online service providers for the safety of their users and instigation of online harms.

[3] Australian Government Department of Communications and the Arts, *Online Safety Legislative Reform: Discussion Paper* (December 2019) 3.

CONSULTATION QUESTIONS

The AHCN addresses questions from the discussion paper below.

High level objectives

Are the proposed high level objects appropriate? Are there any additions or alternatives that are warranted?

It is currently proposed that a new Online Safety Act include a set of high level objects of:

1. preventing online harms;
2. promoting online safety; and
3. protecting Australians online.

The social policy objectives of the proposed Act are too general and need clearer articulation if government is to ensure the Charter, BOSE/COSE, as well as the Act, work together most effectively.

Recommendation 1

The AHCN's recommended objectives for the Online Safety Act are:

a. To protect community confidence in the online sphere as a safe space that reflects values[4] of respect for the freedom and dignity of people, commitment to the rule of law, a spirit of egalitarianism that embraces mutual respect, tolerance, fair play, compassion for those in need, and pursuit of the public good.
b. To achieve parity in protections against criminal and unlawful conduct, whether it occurs online or offline.
c. To prevent and minimise online harms to individuals.
d. To prevent and minimise public harm caused directly or indirectly by online conduct, including to the wider community or segments of the community.

Is the proposed statement of regulatory policy sufficiently broad to address online harms in Australia? Are there aspects of the proposed principles that should be modified or omitted, or are there other principles that should be considered?

[4] Australian society values respect for the freedom and dignity of the individual, freedom of religion, commitment to the rule of law, Parliamentary democracy, equality of men and women and a spirit of egalitarianism that embraces mutual respect, tolerance, fair play and compassion for those in need and pursuit of the public good: Australian Government, Department of Home Affairs, *Australian Values Statement,* <https://immi.homeaffairs.gov.au/help-support/meeting-our-requirements/australian-values>.

The Discussion paper states that for 'the purposes of this consultation process, the proposed statement of regulatory policy would indicate that the Act is seeking to:

- implement practical measures to protect Australians against exposure to illegal and harmful online content, with particular regard to the needs of Australian children;
- articulate clear expectations of the online services sector as to its responsibilities to keep Australians safe online;
- require appropriate accountability, transparency and user safeguards from online services sector;
- provide a safety net for users where the online services sector fails to meet its obligations under the Act;
- provide a responsive and flexible approach to online safety;
- balance the competing objectives of user safety and freedom of expression;
- empower and encourage the online services sector to develop solutions for online safety risks as far as possible; and
- encourage the development and use of new technologies and safe products and services.'

Recommendation 2

The AHCN supports the above regulatory statement with the following amendments:

a. The word "online" be removed at the end of the second clause, acknowledging that Australians can be endangered online **and** offline by online conduct.
b. This new clause be inserted: 'introduce mechanisms to overcome common barriers to justice online, particularly the investigation and prosecution of criminal offences.'

Basic Online Safety Expectations

Is there merit in the BOSE concept?

Continuums are an emerging approach of designing public policy in industries where innovation is integral. For example, in education, national curriculum[5] and teacher development[6]; but also, in the information technology area.[7]

In line with the Australian schools' curriculum, continuums have been created to help map where students are in their development of personal and social capability, intercultural and ethical understanding, ICT, as well as across academic subject areas. Where a student may not be meeting a benchmark for that level, the continuum approach allows the teacher and student to understand how many steps they are from reaching it, and how to incrementally build capability towards that point. Conversely, it allows the teacher and student to check if they have met or surpassed the benchmark, and if so, what is next, fostering a culture of continual improvement. In other spheres, it has been used to optimise processes, including where the sector is capable of lifting the bar but needs to make those changes more systemically and reliably.

Recommendation 3

The AHCN recommends that the Australian Government consult with industry and key community stakeholder groups on the merits of the concept of **Continuum of Online Safety Expectations**, instead of the concept of Basic Online Safety Expectations. The problem with simply setting a minimum standard through BOSE is that it does not foster a culture of continual improvement. A continuum is a preferable tool to demonstrate to an entity where it might exist on a current spectrum of providers, and the appropriate expectations that are attributed to it by virtue of its profit margin or user base size.

Are there matters (other than those canvassed in the Charter) that should be considered for the BOSE? Are there any matters in the Charter that should not be part of the BOSE?

---

5 See, eg, Australian Curriculum Assessment and Reporting Authority, 'Personal and Social Capability Learning Continuum', <https://www.australiancurriculum.edu.au/media/1078/general-capabilities-personal-and-social-capability-learning-continuum.pdf>.

6 See eg, Australian Institute for Teaching and School Leadership, 'Classroom Practice Continuum', *Australian Professional Standards for Teachers,* <https://www.aitsl.edu.au/docs/default-source/default-document-library/looking-at-classroom-practice-continuum---a3-summary.pdf?sfvrsn=146e23c_2>.

7 The Capability Maturity Model's (CMM) aim is to improve existing software development processes, but it can also be applied to other processes.. The term 'maturity' relates to the degree of formality and optimization of processes, from *ad hoc* practices, to formally defined steps, to managed result metrics, to active optimization of the processes: Mark Paulk, Charles Weber, Bill Curtis, Mary Chrissis, 'Capability Maturity Model for Software (Version 1.1)', (Technical Report, Software Engineering Institute, Carnegie Mellon University, February 1993).

The AHCN proposes that the following matters be considered for the COSE/BOSE, and for additional legislative reform, where stated.

Quality Stakeholder Engagement

The COSE/BOSE should recognise the need for engagement between industry, government, community and academic stakeholders to continually improve provider guidelines. Community confidence can only be maintained through this engagement.

For example, Facebook has developed its own hate speech guidelines: [8]

> We only remove content that directly attacks people based on certain protected characteristics. Direct attacks include things such as:
>
> - Violent or dehumanising speech
>   For example, comparing all people of a certain race to insects or animals
> - Statements of inferiority, disgust or contempt
>   For example, suggesting that all people of a certain gender are disgusting
> - Calls for exclusion or segregation
>   For example, saying that people of a certain religion shouldn't be allowed to vote

Exactly how Facebook defines a 'direct attack' as opposed to an 'indirect attack' is yet to be clarified, and could be the reason that Facebook has come under consistent criticism for failing to act on the proliferation of extreme hate rhetoric, including from far right groups, which might tactically enable them to avoid detection.[9]

Recommendation 4

For high achieving or higher capacity[10] entities on the proposed continuum (COSE), there are areas that can be improved and fine-tuned through stakeholder engagement. Stakeholder engagement needs to include academia in the field of hate crime and far right extremism; as well as civil society representing targeted community groups and online hate prevention.[11]

---

8 The text displayed by Facebook when a user attempts to report 'hate speech' in January 2020. See Facebook, *Community Standards* <https://www.facebook.com/communitystandards/>.

9 See, eg, Puecker, Smith, and Iqbal, above n 2, 9: "In the posts, words like Jew, Jewish or Zionist appeared relatively rarely, however they were much proportionally frequent in the comments (+193%). This divergence was initially surprising given that the groups were openly anti-Semitic and that posts usually only attracted a small number of comments. There are at least two possible explanations for this. First, the group leaders were implicitly alluding to anti-Semitic tropes in some of their posts (possibly in an attempt not to breach FB community standards), which followers then made explicit. Second, group leaders simply avoided anti-Semitic messaging (possibly to avoid being shut down by FB), but the followers still brought their anti-Semitic themes into the discussion through comment." See Christopher Knaus and Michael McGowan, 'Far right hate factory still active on Facebook despite pledge to stop it', *The Guardian* (Australia), 5 February 2020. See also Dr Andre Oboler, 'Islamophobia on the internet: The growth of online hate targeting Muslims' (Online Hate Prevention Institute, December 2013). Dr Andre Oboler, 'Anti-Muslim hate still online' (Online Hate Prevention Institute, 2014): In 2013, the Australian based Online Hate Prevention Institute examined and reported 50 Facebook pages for displaying anti-Muslim hatred. Before their report was published, 6 pages were shut down - a year later 10 more were shut down. The vast majority remain active and since then, more have opened.

10 Capacity would have regard to user base size and profit margins.

11 For example, the AHCN notes the important role played by the police funded third party website, True Vision, in the UK, in engaging with social media service providers: See European Union Agency for Fundamental

Recommendation 5

The AHCN also supports the following recommendations from the Australian Taskforce to Combat Terrorist and Extreme Violent Material Online:[12]

a. Relevant Australian Government agencies, academia, researchers, and civil society bodies that monitor and review terrorist and extremist organisations to share with digital platforms (where legally and operationally feasible) indicators of terrorism, terrorist products and depictions of violent crimes. (Rec 4.3)
b. Digital platforms to fund (including via the GIFCT, as appropriate), with the support of the Australian Government, research and academic efforts to better understand, prevent and counter terrorist and extreme violent material online, including both the offline and online impacts of this activity, and use this knowledge to develop and promote positive alternatives and counter-messaging online (Rec 4.2)

Proscription of public advocacy of hatred or prejudice-related violence

It is appropriate that the Australian Government proscribe the public advocacy of hatred or prejudice-related violence against individuals or groups as part of the Online Safety Act and policy instruments.

The limits of section 18C

A civil remedy against racial hatred and offensive behaviour is provided by section 18C of the federal *Racial Discrimination Act 1975* ('RDA').[13] Although this civil remedy was used with great effect with Facebook in 2012[14], the racial hatred provisions of the Act only had 7 online

Rights, *Promising practices: True Vision - the police owned web resource for hate crime,* <https://fra.europa.eu/en/promising-practices/true-vision-police-owned-web-resource-hate-crime>.

12 Comprising government and industry representatives, the objective of the Taskforce was to provide advice to Government on practical, tangible and effective measures and commitments to combat the upload and dissemination of terrorist and extreme violent material. The establishment of the Taskforce was an outcome of a Summit following the Christchurch massacre, on 26 March 2020, which brought together representatives from the major digital platforms, Australian Internet Service Providers (ISPs), the heads of relevant Government agencies, along with the Attorney-General, the then Minister for Communications and the Arts and the Minister for Home Affairs. Civil society, researchers and representatives from targeted communities were not involved in the Summit or subsequent Taskforce, but the Taskforce's final recommendations highlighted the importance of broader engagement: See Australian Government, Prime Minister and Cabinet, *Final Report of the Australian Taskforce to combat terrorist and extreme violent material online* (21 June 2019) <https://www.pmc.gov.au/sites/default/files/publications/combat-terrorism-extreme-violent-material-online.pdf>.

13 It makes it unlawful to do an act, otherwise than in private, that is reasonably likely to offend, insult, humiliate or intimidate a person or group of persons by reason of their race, colour or national or ethnic origin. Section 18D sets out a series of exemptions – academic and artistic works, scientific debate and fair reports or fair comment on matters of public interest are exempt from liability under section 18C if done "reasonably and in good faith".

14 The Executive Council of Australian Jewry (ECAJ), the national peak body representing the Australian Jewish community, brought a complaint under the *Racial Discrimination Act* 1975 (RDA) against Facebook, after it had failed to respond to a number of complaints. After conciliation, Facebook removed 'hundreds of crudely antisemitic racist images and comments that had appeared on 51 Facebook pages': Executive Council of Australian Jewry, Submission No 11 to Parliamentary Joint Committee on Human Rights, *Parliamentary*

complaints in 2018-19, from 97 complaints in total, despite the many instances online.[15] Taking individuals through a complaint process can be an onerous burden on the victim, even more so if requiring the collation, coordination and delivery of evidence on social media platforms such as Facebook. Whilst this is an undoubtedly important avenue for victims, it cannot be left to victims to enforce laws and regulate transnational online social media service providers.

The scope of s18C is also limited in terms of who it protects.

The current national legislative framework does not protect faith groups such as Muslims. Whilst the law considers Jews and Sikhs as a common race of 'ethno-religious' origin, Muslims and other faith groups with diverse ethnic backgrounds, have no effective protections at the national level and in half the state jurisdictions.[16] Research shows that bigotry and prejudice targeting Muslims is often 'racialised'[17] and has the same qualities and effects[18] as racism. According to 2016 research by the e-Safety Commissioner, Muslims were one of the most highly targeted groups of online hatred. [19]

There are also no legislative safeguards from vilification on the basis of gender identity, gender expression, sex and sex characteristics and sexual orientation at a Commonwealth level. The level of protection offered in the states and territories varies.[20]

For the disabled community, protection against vilification only exists in Tasmania and the ACT.[21]

Inquiry into Freedom of Speech (6 December 2016), <https://www.ecaj.org.au/wordpress/wp-content/uploads/2016/12/ECAJ-Submission-to-Parliamentary-Inquiry-into-Freedom-of-Speech-6-December-2016.pdf>.

15 Australian Human Rights Commission, *2018-2019 Complaint Statistics*, <https://www.humanrights.gov.au/sites/default/files/2019-10/AHRC_AR_2018-19_Stats_Tables_%28Final%29.pdf>.

16 Australian Muslim Advocacy Network et al, 'Submission to the Australian Government Attorney General's Department, *Joint Submission on the First Exposure Draft of the Religious Discrimination Bill* (2 October 2019) <www.aman.net.au>. In NSW, Muslims were found to not have a singular 'ethno-religious' origin, and as religion was not a protected attribute under NSW vilification law, the vilifying conduct was deemed lawful: eg, *Ekermawi v Nine Network Australia Pty Limited* [2019] NSWCATAD 29.

17 For eg, in reports to the *Islamophobia Register Australia* in 2019, Muslims were referred to as dirty Islamicinese, invasive species, sand niggers. See Online Hate Prevention Institute, Submission to the Legal and Social Issues Committee Legislative Assembly, Parliament of Victoria, *Inquiry into Anti-vilification protections* (14 January 2020) 4. Gail Mason, Natalie Czapski, 'Regulating Cyber-Racism' (2017) 41 *Melbourne University Law Review* 284, 290; Margaret Thornton and Trish Luker, 'The Spectral Ground: Religious Belief Discrimination' (2009) 9 *Macquarie Law Journal* 71, 74-76.

18 An academic analysis of verified hate incidents reported to the Islamophobia Register Australia has been ongoing since 2014, including the short term and long-term impacts on victims: Dr Derya Iner (ed),' Islamophobia in Australia Report II 2017-2018' (Sydney: Charles Sturt University and ISRA, 2019); See also: Dr Derya Iner (ed), 'Islamophobia in Australia 2014-2016' (Sydney: Charles Sturt University and ISRA, 2017).

19 53% of Australian youth aged 12-17 years old had witnessed harmful hate content targeting Muslims: eSafety collaborated with the Department of Education and Training (DET) in this research, which included a national online survey of 2,448 young people aged 12 to 17 conducted between 25 November and 14 December 2016: Australian Government, e-Safety Commissioner, *Young People and Social Cohesion* (2016).

20 Gabi Rosenstreich,' LGBTI People: Mental Health & Suicide Briefing Paper', *National LGBTI Health Alliance,* Revised 2nd Edition, 2013 <https://www.beyondblue.org.au/docs/default-source/default-document-library/bw0258-lgbti-mental-health-and-suicide-2013-2nd-edition.pdf?sfvrsn=2>.

21 See Schedule 3.

An appropriate threshold

In the context of the Online Safety Act and policy instruments of the Charter and BOSE/COSE, consideration needs to be given to a standard that can be effectively enforced by the e-Safety Commissioner.

All state jurisdictions have some sort of civil vilification legislation with the exception of Western Australia and the Northern Territory. In these statutes, the threshold is usually set at '[inciting] hatred towards, serious contempt for, or severe ridicule of', and in two states, 'revulsion' is also included. Civil vilification law operates through a complaint, conciliation and judgement process of administrative law, where there is judicial oversight of where the line is drawn, which allows a lower threshold to be set. For the purposes of the Online Safety Act, the threshold is applied by the e-Safety Commissioner, not a tribunal or court, and needs to be reserved for cases that are clearly harmful. To be a useful tool in the fight against online incitement, the Commissioner must be able to interpret and apply this standard with confidence. However, setting the threshold too high, at the standard of incitement to violence, will also result in inaction on a vast volume of hate material targeting specific groups, where violence may be implied, or different harms such as discrimination or hostility are incited.

Hence, the AHCN proposes that consideration be given to a different legal threshold that is consistent with international standards. This is a threshold that sits above state vilification laws, but below incitement to violence, providing a modest, but potentially effective counterpoint to the fomentation of hatred online. We discuss this threshold further below.

Recommendation 6

a. The AHCN proposes that the Australian Government consider the merits of an anti-incitement legislated standard that offers protection to all groups with established vulnerability in regard to hate crime.

b. One option is to use the international standard contained in the ICCPR[22] - a standard that would aim to minimise the private harms to constituent members of those groups, and public harms to social cohesion, public health and safety.[23]

c. The explanatory memorandum in any such legislation should recognise that extreme examples of hate speech normalise and encourage public acts of hatred and violence against people offline and should not be tolerated in the online sphere. It should also explain the basis of any relevant exceptions (see below). The AHCN is mindful of setting a sufficiently high threshold for effective enforcement by the e-Safety Commissioner and guarding against unnecessary encroachments on freedom of expression.

---

22 *International Covenant on Civil and Political Rights*, opened for signature 19 December 1966, 999 UNTS 171, 6 ILM 368 (entered into force 23 March 1976). As at 2019, there were 167 states parties. See in particular, Art 20(2).

23 The private and public harms of online hatred are summarised later in this document.

d. To move this idea forward, recognising that such a standard would need to be thoroughly researched, debated and tested with community stakeholders, the AHCN has crafted possible wording to demonstrate the concept, drawing from the ICCPR standard as a basis:

    (i) A person must not, on the ground of a protected attribute of another person or class of persons, publicly advocate hatred, in a way that constitutes incitement to discrimination, hostility or violence, towards that other person or class of persons.[24]

    Note:
    "Protected attributes" includes age, ethnic/national origin, disability, homelessness, race, religious belief or affiliation; gender identity;[25] *intersex status; sexual orientation, and HIV/AIDS status.*
    "publicly advocate" includes use of the internet or e-mail to publish or transmit statements or other material.

    (ii) For the purposes of sub-section (i), advocate may be constituted by a single occasion or by a number of occasions over a period of time.

e. The ICCPR also provides that freedom of expression should only be limited where it is provided by law and "necessary for respect of the rights or reputations of others, for the protection of national security, public order, or public health or morals." 26

f. The determination of what this means in the Australian legislative context should have regard for existing exceptions within state vilification legislation. There is substantial alignment and consistency between these exceptions, such as excluding 'fair reports' of the conduct of another person; material that would be subject to absolute privilege in defamation proceedings and 'a reasonable act, done in good faith, 27 for academic, artistic, scientific or research purposes or for other purposes in the public interest (including reasonable public discussion, debate or expositions.' 28 Some exceptions are less settled. For example, expression for a genuine religious purpose is specified within exceptions to the Victorian legislation, and parts of the NSW legislation. 29

---

24 ICCPR, above n 22.

25 See Schedule 2 for examples of the protective scope of hate crimes.

26 ICCPR, above n 22, arts 19 (2), (3).

27 The terms 'reasonably and in good faith' provide some additional parameters. Some considerations may include but are not limited to whether the conduct was honest; the motives behind the conduct; whether there was any malice or careless disregard for the truth from the respondent: *Bropho v Human Rights and Equal Opportunity Commission (*2004) 135 FCR 365. Conduct was found to be unreasonable that involves 'highly inflammatory language', 'offensive imputations', or contains content that is 'so ill-informed, misconceived, ignorant, so hurtful, that it goes beyond the bounds of what tolerance should accommodate': *Menzies and Anor v Owen* [2014] QCAT 661.

28 *Civil Liability Act 1936 (SA)* s73.

29 Schedule 3 to this submission lists the thresholds and exceptions of state and territory jurisdictions. In 2006, the Victorian legislation was amended to clarify that 'religious purpose' included, but was not limited to, *conveying or teaching* a religion or proselytising. 'Religious instruction' is also recognised within defences to

g. The AHCN recommends careful consideration of which exceptions would be necessary and justified in this context, noting the higher threshold of the proposed wording.

h. In examining the merits of such a provision, consideration should be given to the needs for legislative reform in the offline sphere as well. Nonetheless, it would be well placed in the Online Safety Act to protect from harmful online behaviour.

i. This legislated standard would be acknowledged in the Online Safety Charter of community expectations and embedded within the operational accountability requirements of the COSE/BOSE, as a pain point and minimum expectation for platforms and online services. A COSE must continue to reward innovation and best practice in relation to creating safe spaces free from prejudice-motivated online hatred.

Reducing reliance on user complaints and appeals

While agreeing to community standards is important, the failure to effectively implement those standards is harmful. Later in this submission, the AHCN recommends that the public advocacy of hatred and violence ought to be brought within the ambit of e-Safety Commissioner's redress scheme.

While we support platform-based appeal mechanisms being part of the COSE/BOSE, when it comes to the public advocacy of hatred and violence, the volume of content is prolific and toxic. For victims from targeted communities to exhaust reporting and appeal mechanisms for each post is onerous. The weakness of this system can also be exploited by offenders who use bots and other technologies to amplify their advocacy for hatred and violence.

Recommendation 7

As a result, the framework offered by COSE/BOSE and the proposed Act needs to work together to incentivise social media platforms to self-moderate more effectively, recognising that reliance on user-reporting mechanisms is not appropriate in this context.

Appropriate handling of repeat offenders

Vital to community confidence is how social media companies handle repeat offenders.

Recommendation 8

The AHCN advocates for consideration to be given to recent recommendations from the Online Hate Prevention Institute that: [30]

homosexual vilification; and 'religious discussion and instruction' within defences to transgender and HIV/AIDS vilification law in NSW.

[30] See detailed recommendations for industry and government: Dr Andre Oboler, Dr Patrick Scolyer-Gray and William Allington, 'Hate and violent extremism from an online sub-culture: The Yom Kippur Terrorist Attack in Halle, Germany' (Online Hate Prevention Institute, December 2019), Recommendations 20-24.

a. Facebook and other social media companies need to start actively cooperating with governments on lower level incidents. There needs to be real consequences when pages and accounts are closed for hate speech. Account suspensions should lead to account closures when harmful behaviour continues.
b. When a page is closed, the page administrators should lose the right to be administrators of other pages for a period of no less than 12 months. A second page being banned should see the ability to administer pages permanently blocked. With time, Facebook may wish to require page administrators to be verified based on government issued identification documents. This would help to prevent banned administrators opening additional accounts under false or alternate aliases.

Cooperation with removing barriers to justice

Recommendation 9

The AHCN also supports the Online Hate Prevention Institute's recommendation for legislation requiring a platform to provide **a tool for rapid verification of jurisdiction[31]:**

> Jurisdiction is one of the challenges that complicates online regulation. It is particularly acute with online platforms which store and then pass on a user's communications. To address this, we believe it is necessary for legislation to require platforms to provide a tool to verify if a user is within the State's jurisdiction. Such a tool should work in real time giving an immediate 'yes or no' to the question of jurisdiction. For example, whether a user is in Victoria or Australia.

Experience has shown when it comes to cyber investigations, although IP addresses identify a location, it is next to impossible to prove who sent the message.

Recommendation 10

One solution is for government to consider the merits of legislation which allows regulators to place a **'form of demand'** on online users alleged to have committed an offence. A legislative precedent for this proposal can be found in section 177 of the *Roads Transport Act 2013* (NSW) and section 14 *Law Enforcement (Powers and Responsibilities) Act 2002* (NSW) whereby police have the power to place a form of demand in relation to vehicles involved in offences. In the same way the NSW legislation allows the police to identify the driver of a vehicle at the time of the offence, the introduction of analogical legislation on the online space would require the person responsible for the IP address to identify the 'user' responsible for online crimes. In this analogy, the 'motor vehicle' subject of an online investigation is the 'user account'.

Accountability for spreading false information and failing to remedy

31 Online Hate Prevention Institute, above n 17,7.

Giant social media companies are hosting false news stories that perpetuate prejudice-motivated hatred against segments of the community.[32] Those platforms are not doing enough to discharge their responsibilities to prevent the spread of false news when identified, nor enacting consequences for offending account holders. Only where there is significant public and political pressure, do platforms appear to take more substantial or innovative action.[33]
In November 2019, Australia's Attorney-General the Hon. Christian Porter said *"[t]he playing field between digital platforms and mainstream media is completely uneven"* and that *"[m]y own view is that online platforms, so far as reasonably possible, should be held to essentially the same standards as other publishers."*[34] These comments were made in relation to defamation law reform, but in terms of consistency of law, it would make sense that such reform be considered for other substantial breaches of 'traditional' media standards.

Recommendation 11

Accordingly, the AHCN recommends:

a. A new category of liability be considered for online services to be held accountable for publishing inaccurate or misleading information[35] that perpetuates hatred or hostility towards a segment of the community; and for failing to correct or take adequate remedial action if published material is significantly inaccurate or misleading[36].
b. Pages that repeatedly publish clearly false news stories that perpetuate hate and prejudice must face appropriate consequences (as recommended above 'handling of repeat offenders')
c. Expanding on best practice and accountability in this area should be included within the remit of COSE/BOSE.

Improving platform moderation processes in relation to extremist content

Recommendation 12

The AHCN recommends that the following matters be included as part of the performance evaluation of online services in the COSE/BOSE. These are selected recommendations from the Australian Government's own Taskforce review, following Christchurch[37]:

---

32 An American study investigated the differential diffusion of all of the verified true and false news stories distributed on Twitter from 2006 to 2017. The data comprise ~126,000 stories tweeted by ~3 million people more than 4.5 million times. Falsehood diffused significantly farther, faster, deeper, and more broadly than the truth in all categories of information. The degree of novelty and the emotional reactions of recipients may be responsible for the differences observed: Soroush Vosoughi, Deb Roy and Sinan Ara, 'The spread of true and false news online', *Science*, 9 March 2018, <https://science.sciencemag.org/content/359/6380/1146>. See also Alice Marwick and Rebecca Lewis, 'Media Manipulation and Disinformation Online' (New York, Data & Society Research Institute 2017), <https://datasociety.net/pubs/oh/DataAndSociety_MediaManipulationAndDisinformationOnline.pdf>.

33 See Michelle Toh, 'Facebook, Google and Twitter crack down on fake coronavirus 'cures' and other misinformation' *CNN news* (Hong Kong) 3 Feb 2020. Dave Lee, 'Matter of Fact Checkers: Is Facebook winning the Fake News war?' *BBC News* (North America) 2 April 2019.

34 Fergus Hunter, 'Law should treat social media companies as publishers: Attorney-General', *Sydney Morning Herald* (Sydney) 20 November 2019.

35 Australian Press Council, *Statement of General Principles.* <https://www.presscouncil.org.au/statements-of-principles/>

36 bid.

37 Australian Government, Prime Minister and Cabinet, above n 12.

a. Digital platforms to review the operation of algorithms and other processes that may drive users towards (or amplify) terrorist and extreme violent material to better understand possible intervention points, and to implement changes where this occurs. This may include using algorithms and other processes to redirect users from such content, or the promotion of credible, positive alternatives or counter-narratives (Rec 1.4).
b. Digital platforms to fund (including via the GIFCT, as appropriate), with the support of the Australian Government, research and academic efforts to better understand, prevent and counter terrorist and extreme violent material online, including both the offline and online impacts of this activity, and use this knowledge to develop and promote positive alternatives and counter-messaging online (Rec 4.2)

What factors should be considered by the eSafety Commissioner in determining particular entities that are required to adhere to transparency reporting requirements (e.g. size, number of Australian users, history of upheld complaints)?

The AHCN supports the Government's election commitment to mandate transparency reports from major social media platforms that provide data on the number and type of responses to reports and complaints about illegal, abusive and predatory content by users.[38]

Recommendation 13

The AHCN makes the following recommendations:

a. The legal mechanisms used to mandate, or request transparency reports need clarification.
b. The history of upheld complaints, or published reports from research institutions or civil society[39], which provide evidence of harmful content being published on a particular platform, should be able to trigger a request for a transparency report.
c. 'Illegal, abusive and predatory content' must include material that publicly advocates hatred or violence against segments of the community, including the promulgation of extremist and/or biased narratives about segments of the community.[40]
d. This mechanism should be appropriately linked to the Continuum of Online Expectations (COSE/BOSE). An entity that refuses to provide a transparency report or provides partial data should become liable to face sanctions proportionate to its national and international revenue. This is a point that can be articulated in the COSE/BOSE.

---

38 Australian Government, above n 3.

39 See, eg, Iner, above n 18; Oboler, above n 9 (Facebook); Oboler, above n 30 (4chan and 8chan). Each year a report is compiled by ECAJ Research Director Julie Nathan, covering antisemitic incidents and antisemitic discourse in Australia: Julie Nathan, *Report on Antisemitism in Australia Report 2019,* (Executive Council of Australian Jewry, 24 November 2019), <http://www.ecaj.org.au/2019/the-ecaj-2019-antisemitism-report/>.

40 See, eg, Puecker, Smith and Iqbal, above n 2. See also Australian Government, Prime Minister and Cabinet, above n 12 :Relevant Australian Government agencies, academia, researchers, and civil society bodies that monitor and review terrorist and extremist organisations to share with digital platforms (where legally and operationally feasible) indicators of terrorism, terrorist products and depictions of violent crimes (Rec 4.3)

Should there be sanctions for companies that fail to meet the BOSE, beyond the proposed reporting and publication arrangements?

Consideration of a new category for social media companies with respect to media laws

The Government's discussion paper canvasses a range of regulatory responses implemented in Germany, France, Europe more generally and Canada.

Recommendation 14

The AHCN supports consideration of the recommendation of the UK White Paper on Online Harms that:

> Social media companies cannot hide behind the claim of being merely a 'platform' and maintain that they have no responsibility themselves in regulating the content of their sites. We repeat the recommendation from our Interim Report that a new category of tech company is formulated, which tightens tech companies' liabilities, and which is not necessarily either a 'platform' or a 'publisher'. This approach would see the tech companies assume legal liability for content identified as harmful after it has been posted by users. We ask the Government to consider this new category of tech company in its forthcoming White Paper.[41]

Appropriate consequences aligned to expectations continuum

The Online Hate Prevention Institute based in Australia, writes:

> Major platforms should be under an obligation to prevent the display of material which is unlawful [in Australia] to users of their platform who are connecting from [Australia]. We see no reason why the same standard and time frames could not apply here as applies in Germany.[42]

Recommendation 15

The AHCN agrees it is reasonable to penalise platforms for failing to comply to a reasonable standard. As explained above, the AHCN submits it would be reasonable to align expectations and consequences within a Continuum of Online Safety Expectations (COSE). This means that social media service providers at the top of the continuum who have the financial power to achieve a high level of safety but fail to do so, would face penalties. Platforms that continually allow criminal breaches or refuse to moderate or report, and exist at the lowest end of the online safety continuum, should face the risk of being deplatformed.

Liability of online social media service providers for hosting criminal content

Recommendation 16

If an online media service provider is found to have hosted illegal and harmful content, and not acted against the distribution of this content, the e-Safety Commissioner should have the ability

---

41 UK Department for Digital, Culture, Media and Sport, *Online Harms White Paper*, 8 April 2019, para 14 < https://www.gov.uk/government/consultations/online-harms-white-paper/online-harms-white-paper>.

42 Online Hate Prevention Institute, above n 17, 8. See also Oboler, above n 30, Recommendation 38.

to launch legal proceedings against them, with the prospect of large fines being administered. The failure to adequately respond to repeat offenders should be treated as an aggravating factor. These fines can be reduced if an online media service provider implements measures to improve its performance in line with the COSE.

Cyberbullying and Cyberabuse

The AHCN supports the extension of powers to issue take down notices in cases of cyber abuse affecting adult victims. It makes sense to extend the cyberbullying/ cyberabuse scheme for children and adults to designated internet services and hosting services, relevant electronic services and social media services. In addition, the 24 hours take down period (instead of 48 hours) reduces the amount of potential harm to victims. However, the AHCN has some further recommendations:

Recommendation 17

a. The system should find a way to acknowledge those providers with the fastest take-down rates, to encourage continual improvement. This could be incorporated into the Continuum of Safety Expectations (COSE/BOSE), with high achieving platforms and services receiving a star rating that they can advertise with potential users, especially families looking for child-safe platforms.
b. The AHCN also advises that if cyber-bullying or cyber abuse has a dimension of prejudice, that dimension needs to be recognised and treated as an aggravating factor. The online service or platform must respond accordingly, by implementing whatever standard or protocol they have in relation to hate speech, as well as bullying. This is essential from a data integrity perspective, but also in terms of restorative justice and healing for the victim. Where bullying incidents involve a hate or prejudice element, based on a protected attribute of the victim, and that element is overlooked or ignored, this tends to intensify the distress of the victim. Both elements need to be identified, labelled and responded to.
c. The AHCN recommends further consultation with targeted communities to understand specific harmful, online practices and speech with regard to their context. For example, in terms of online safety for LGBTIQ+ people, one of the most dangerous and harmful practices is “outing” someone. This is not always detected as bullying, abuse, or even ‘hate speech’, but can result in violence, intrafamilial hate crime, and in some extreme circumstances, homicide (see for example the recent murders in Russia stemming from a campaign of outing LGBTIQ+ people).[43] Similarly, for disabled people, speech often deployed in argument – such as stupid, idiot, moron – is considered ableist hate speech by some disabled people. Additionally, in terms of online safety, for some disabled people (with cognitive impairments), the main issue is the manipulation of their vulnerability – often from “mates” (see the term “mate crime”).[44] This also relates to the experiences of older people when elder abuse presents as hate crime. Without further

---

[43] Tim Fitzsimons, ‘Russian LGTBQ activist is killed after being listed on gay-hunting website’ *NBC News* (USA) 24 July 2019.

[44] See Pam Thomas, ‘Mate Crime: Ridicule, Hostility, and Targeted Attacks against Disabled people’ (2011) 26(1) *Disability and Society* 107; Endeavour Foundation, *Preventing Mate Crime: For people with an intellectual disability* (1 December 2016) < https://www.endeavour.com.au/media-news/blog/mate-crime>. See eg, Peter Walker, ‘Gemma Hayter case review finds chances were missed to protect her’, *The Guardian News* (UK) 15 November 2011.

engagement with other targeted groups, these unique and poorly understood forms of online abuse and harassment may be missed in the operationalisation of the Online Safety Act.

d. The AHCN is interested in what recourse a victim will have if the e-Safety Commissioner refuses to take down material.

Addressing illegal and harmful content

The current legislative framework identifies "extremely violent content, terrorist propaganda, or child sexual abuse and exploitation material" as illegal and harmful content.[45]

G20 leaders made a statement at the Osaka Summit that:

> "The internet must not be a safe haven for terrorists to recruit, incite or prepare terrorist acts. To this end, we urge online platforms to adhere to the core principle, as affirmed in Hamburg, that the rule of law applies online as it does offline."

Existing criminal offences relating to the incitement of hatred or violence are almost never applied in the online space, due to the potential inadequacy or inappropriateness of those laws, but also due to a lack of police training, policy focus from relevant agencies (including dedicated hate crime units), and difficulty in handling jurisdiction.[46] The AHCN lists these criminal laws in Schedule 1. Earlier in this submission, the AHCN made some recommendations about removing barriers to justice for potential crimes conducted online.
The AHCN has also proposed criminal law reform in relation to hate crimes, including criminal vilification and incitement to violence.[47] However, there is also a role for the e-Safety Commissioner in ensuring clear examples are removed by platforms as soon as practicable.

Extending the Commissioner's Take Down powers

Recognising the challenges in policing the public advocacy of hatred and violence, in terms of balancing community/user safety and freedom of expression, the AHCN acknowledges that the Commissioner's powers should be evoked at a sufficiently high threshold; and online platforms and services need to be fully engaged in a process of continual improvement.

Abhorrent materials such as child pornography and exploitation material, terrorist propaganda, and extreme violence are included because of the degree of societal harm they pose.

The compounding, echo-chamber effect of social media, overtime, can erode Australian values and moral social norms against abusing someone based on certain characteristics. As a result of abuse online and offline, victims often withdraw and disengage, emboldening the hate group and enabling their growth to continue unchallenged.[48] Eventually, the social stigma against prejudice, discrimination and hatred ceases to exist,[49] and examples of online hatred can

---

45 Australian Government, above n 3, 12.

46 The AHCN details many of the issues in relation to the NSW criminal justice system and hate crime, including policing issues, and discusses options for criminal law reform in its submission to the *NSW Parliamentary Inquiry into Gay and Transgender hate crimes between 1970 and 2010*. The submission will be published on the AHCN website: <https://sydney.edu.au/law/our-research/research-centres-and-institutes/australian-hate-crime-network.html>.

47 Ibid.

48 Mason and Czapski, above n 17, 295-96.

49 Dr Andre Oboler, 'Legal Doctrines Applied to Online Hate Speech' (2014) 4 *ANZ Computer Law Journal* 9, 11. See eg, the *Islamophobia in Australia Report 2019* found 'Contrary to expectation, the majority of reported

graduate into organised and opportunistic violence offline. Significant dangers arise when public advocacy of hatred becomes mainstream.

Currently, as Mason and Czapski write, 'there is no comprehensive system for expressly denouncing and remedying the harm of cyber-racism [or other forms of online hatred] by offering an efficient and accountable process for removing harmful material, backed by a mechanism of enforcement.'[50]

Given the variety and complexity of criminal laws in this space (see Schedule 1), it may be administratively difficult for the Commissioner to assess whether content should be removed using these laws as the threshold. A number of these laws also operate only at the point of incitement to violence. This is too high a threshold for the purposes of achieving the higher objectives of the Online Safety Act. The standard the AHCN has highlighted in Recommendation 6 is more appropriate in the civil jurisdiction (which does not involve criminal sanctions) as it combines the public advocacy of hatred with the incitement of either discrimination, hostility or violence.

Recommendation 18

a. In these circumstances, for the purposes of establishing a threshold, the e-Safety Commissioner could either:

   i. Use the 'civil standard' proposed earlier in this document as a threshold to issue take down notices (Recommendation 6); or
   ii. Extend the definition of cyber-abuse to cover abuse expressed to whole groups of people on the basis of their identity.

b. Where there are multiple requests for take down notices against the same platform, further sanctions ought to be enlivened in accordance with the COSE/BOSE.

incidents (60%) occurred in guarded or patrolled areas, where police officers, security guards, track-work personnel, and other workers or officials were in force, or surveillance cameras deployed. The increasing harassment in guarded places since the previous report (30%) is an alarming security problem': Iner, above n 18, 6.

50 Above n 17, 287.

Blocking measures for terrorist or extremist violent content

Ancillary service provider notice scheme

The AHCN sees merit in measures to block terrorist or extremist violent content, as well as the concept of an ancillary service provider notice scheme.

The AHCN agrees there is merit to making compliance with the ancillary service provider notices mandatory.

The scheme activates in situations where third parties 'systematically and repeatedly' facilitate the posting of cyberbullying or cyber abuse material, image-based abuse or hosting illegal or harmful content.

Recommendation 19

The AHCN submits that use of the phrase 'systematically' would present significant difficulties for regulators and prosecutors.

Referral of extremist manifestos for classification

The office of the e-Safety Commissioner does have remit in relation to extremist propaganda and referring manifestos to the Classification Board. However, it has been reported that it will only do so if the Office considers the manifestos to be at risk of going viral in Australia.[51] The narratives of those manifestos are freely shared online without consequence from the Commissioner's Office, police or social media service providers.

Recommendation 20

a. The e-Safety Commissioner needs to ensure that extremist ideologies that incite, inspire or influence terror attacks result in RC classification.

b. The AHCN asks the Government to construct appeal mechanisms should the e-Safety Commissioner fail to exercise this power.

Challenging extremist ideologies

While the AHCN sees merit in the above measures, the AHCN believes that further steps are needed to mitigate the expansion of far right extremism online. With no far right organisations listed as proscribed terrorist organisations, and limited legislative guidance on a national level as regards to public advocacy of hatred, there are few policy or legislative incentives for providers to take action against the mobilisation frames used by these extremists.[52]

---

51 Cameron Wilson, 'Australia Won't Ban Manifestos Similar to The Christchurch Shooter's Because They Didn't Go as Viral', *Buzzfeed news* (Sydney), 16 January 2020.

52 See Puecker, Smith and Iqbal, above n 2.

According to the GIFCT website, it plans to 'publish a cross-platform, countering violent extremism toolkit, developed with the Institute for Strategic Dialogue, to help civil society organisations build online campaigns that challenge extremist ideologies, while prioritising safety.'[53]

Recommendation 21

The AHCN seeks further information and engagement in the development of the GIFCT toolkit. To the extent civil society is conferred this role, it will require resourcing to complete the task.

53 Global Internet Forum to Counter Terrorism, *Next Steps for GIFCT,* 23 September 2019, <https://gifct.org/press/next-steps-gifct/>.

Role of the e-Safety Commissioner

Below are two further functions where the Commissioner's role could be expanded.

Making justice more achievable in the online space

Recommendation 22

The AHCN submits that an objective of the Commissioner's Office should include making the law easier to apply in the online space in order to maintain community confidence both online and offline and facilitate justice more broadly.

A review of programs that reduce recidivism of repeat offenders of online hatred[54]

International research suggests offenders tend to be young males, including adults, mostly from 'white' or majority ethnic groups, unemployed or in poorly paid and insecure jobs.[55] In Australia, there are no rehabilitation programs geared towards these hate crime offenders.[56]

Some solutions to address these gaps include:

- Restorative justice approaches that bring together offenders, victims, criminal justice agents and the wider community to repair the harm, without always resorting to punitive measures such as imprisonment[57]
- Advocacy services that work with hate crime victims to assist in practical matters such as reporting to police as well as therapeutic counselling and safety planning. Examples include Safe Horizon in New York and Victim Support in England and Wales
- Dedicated programs for hate crime offenders to address the motivations for online hate and hate crime by building pro- social attitudes and behaviours. These programs are often directed towards offenders linked to far-right extremism. Examples include EXIT in Sweden and ADAPT in England. Research shows the Swedish program is particularly successful.[58] However, many hate crime offenders are driven by different biases and ideologies, such as those who target the LGBTI or disabled communities. Dedicated programs that are developed to address these offenders are equally important.

---

54 The following section was reproduced from: Professor Gail Mason and Associate Professor Nicole Asquith, 'Islamophobia within a Hate Crime Network' in Dr Derya Iner (ed), *Islamophobia in Australia Report II 2016-2017* (Charles Sturt University and ISRA, 2019) 25.

55 Paul Iganski, David Smith, Liz Dixon, Vicky Kielinger, Gail Mason, Jack McDevitt, Barbara Perry and Andy Stelman, 'Rehabilitation of Hate Crime Offenders' *(*Summary Report, Scottish Equality and Human Rights Commission, 2011a) cited in Ibid.

56 Paul Iganski, David Smith, Liz Dixon, Vicky Kielinger, Gail Mason, Jack McDevitt, Barbara Perry and Andy Stelman, 'Rehabilitation of Hate Crime Offenders' (Research Report, Scottish Equality and Human Rights Commission, 2011b) cited in Ibid.

57 Mark Walters, *Hate Crime and Restorative Justice: Exploring Causes, Repairing Harms*. (Oxford University Press, 2014).

58 Iganski et al, above n 56.

Recommendation 23

The AHCN recommends the e-Safety commissioner partner with researchers to conduct a review of the effectiveness of these rehabilitative and restorative programs used overseas and their mechanisms for identifying participants.

BACKGROUND

Who we are

The Australian Hate Crime Network (AHCN) is a national partnership composed of three sectors of society: academics, representatives of NGOs from minority communities, and people from relevant government departments.

The AHCN aims to:

- provide leadership, advocacy and support for state and national government responses to hate crime and hate incidents;
- provide an educative and advisory role to key agencies and services on preventing and responding to hate crime and hate incidents;
- enhance community awareness of hate crime and hate incidents, and
- encourage reporting, help seeking and access to available resources;
- monitor and review patterns in hate crime and hate incidents;
- advocate for improvement in data collection, law enforcement and criminal justice responses; and,
- collect and distribute relevant current research and knowledge on hate crime and hate incidents.

For more information on the AHCN, visit our website.

Definition of hate crime[59]

Hate crime, also called bias crime, is used to describe criminal and sub-criminal incidents motivated by bias or hatred towards a group of people.

Victims are targeted because of their age, disability, homelessness, race, religious belief or affiliation, gender or gender identity, intersex status, HIV/AIDS status, or sexual orientation.

Not all hate crimes are a breach of criminal law. However, the term is commonly used to capture sub-criminal incidents, such as hate speech, that pave the way for more aggressive or physically harmful behaviour. These incidents are the tip of the iceberg.

Social media platforms such as Facebook have internal guidelines regarding content and Community Standards that reject hate crime and hate speech. Nonetheless, their processes for

[59] The following section was reproduced from: Professor Gail Mason and Associate Professor Nicole Asquith, 'Islamophobia within a Hate Crime Network' in Dr Derya Iner (ed), *Islamophobia in Australia Report II 2016-2017* (Charles Sturt University and ISRA, 2019) 18.

deciding what should be allowed and enforcing these standards continually fail to provide a safe and respectful environment.[60]

Online hatred

In Australia, findings on people's attitudes, awareness and responses to online hate speech come from a nationally representative survey of 3,737 adults aged 18 to 65 about online safety commissioned by eSafety in August 2019.[61] Religion, political views, race and gender were the most common reasons cited in both Australia and New Zealand for experiencing hate speech. Seven in 10 adult Australians believe that online hate speech is spreading with the majority agreeing that more should be done to stop its growth either through the introduction of new laws (71%) or through social media companies doing more (78%).[62]

A 2016 survey by the e-Safety Commissioner[63] of 2448 young Australians aged 12-17 years old found their experiences included:

- Seeing racist comments: 56% - 60% of girls, 53% of boys
- Seeing or hearing hateful comments about cultural or religious groups: 53% - 57% of girls, 50% of boys

Targets of harmful content online (multiple responses allowed):

- Muslims, 53%
- Asylum seekers, 37%
- Aboriginal and Torres Strait Islander peoples, 37%
- Refugees, 35%
- Asians, 33%
- LGBTI, 26%
- Africans, 20%
- Jews, 17%
- Christians, 15%
- Other minority groups, 2%

There was a resurgence of abuse towards the lesbian, gay, bi, trans, intersex (LGBTI) community surrounding the 2017 survey on the legalisation of same-sex marriage.[64]

---

60 Mason and Czapski, above n 17.

61 e-Safety Commissioner, above n 1.

62 Also, a sizeable minority of adult Australians (23%) believe that people should be free to say whatever they want online. This group was predominantly male and those aged 30–40. Moreover, they overwhelmingly identified as being heterosexual, from non-Aboriginal and Torres Strait Islander (Indigenous) and non-culturally and linguistically diverse backgrounds (CALD): e-Safety Commissioner, above n 1.

63 e-Safety Commissioner, above n 19.

64 Stefano Verrelli, Fiona A. White, Lauren J. Harvey and Michael R. Pulciani, 'Minority Stress, Social Support, and the Mental Health of Lesbian, Gay, and Bisexual Australians during Australia's Marriage Law Postal Survey' (2019) 54 *Australian Psychologist* 4, 336-346.

Charles Sturt University's Islamophobia in Australia Report found:

> The most severe level of hate, i.e. wanting to kill/harm Muslims, was the most dominant rhetoric, consisting of the one-quarter of the entire online cases.[65]

The latest ECAJ Report on Antisemitism in Australia documented extensive and extreme examples of public advocacy of hatred and incitement of violence, including mass murder, against Jewish people recorded on websites, Facebook, Twitter, Youtube and on the fringe social media platform, Gab.[66]

Private harm

There is significant and established evidence over many years of the psychological effects of racism, including poorer mental health, reduced quality of life.[67]

Discrimination and social exclusion contribute to LGBTI people experiencing a higher prevalence of other risk factors associated with mental ill-health and suicidality than the rest of the population.[68]

Charles Sturt University's Islamophobia in Australia report highlighted both short and long term health trauma of hate incidents on victims, especially when bystanders did not assist.[69]

As Waldon writes, the harm of hate speech is that it removes victims *'assurance that there will be no need to face hostility, violence, discrimination, or exclusion by others'* in their daily lives.[70]

In their submission to the Parliamentary Inquiry into Freedom of Speech in 2016, the Executive Council of Australian Jewry highlighted the impacts of prejudice (in this case, racism) on the social determinants of health:

- Shaming and degrading a group of people by labelling them inferior ('stigmatising') can inflict psychological injury by assaulting self-respect and dignity.
- Self-esteem and the respect of others are important for participation in society.
- The desire to avoid being continually confronted with speech of this nature, or by actual or potential perpetrators, places limits on the target's freedom to maintain broad support networks, limiting social harmony and circumscribing possibilities to form and maintain personal relationships.[71]

---

65 Iner, above n 18.
66 Nathan, above n 39.
67 Kevin Dunn, Yin Paradies and Rosalie Atie, *Preliminary Result: Cyber Racism and Community Resilience the Survey* (Research Report, 28-29 May 2014, p 3. Victorian Health Promotion Foundation, *Mental health impacts of racial discrimination in Victorian culturally and linguistically diverse communities, Experiences of Racism survey: a summary,* December 2012.
68 Rosenstreich, above n 20.
69 Iner, above n 18, 92.
70 Jeremy Waldron, The Harm in Hate Speech (Harvard University Press, 2012) 2—3 quoted in Dr Andre Oboler, 'Legal Doctrines Applied to Online Hate Speech' (2014) 4 *ANZ Computer Law Journal* 9.
71 Above n 12, 8.

The targets of hatred often limit their own speech as a 'protective measure'[72], and in some cases withdraw physically from public spaces and public facing work.[73]

Public harm

The diversity of Australia's people is a great source of our nation's strength and imposes an obligation on the government to protect and encourage social cohesion.

The societal harm caused by a hate crime is magnified by the sense of fear and danger that permeates through the victim's community group following an attack.

Contemporary surveys of Australian LGBTIQ people show unacceptably high rates of violent victimisation, with reports at alarming levels for transgender people.[74]

Hate speech is also closely linked to severe forms of violent extremism[75], such as the massacre of Muslim New Zealanders in Christchurch on 15 March 2019.

The Online Hate Prevention Institute's report published in December 2019, noted how the /pol/community found on places like 4chan and 8chan have been responsible for four deadly terrorist attacks in 2019.[76]

Victoria University research has identified an 'increasingly radical milieu' that is leaning closer to violent outcomes. A Victorian analysis of far right extremist activity online by Victoria University, published its preliminary findings in November 2018.[77] It studied over 41, 000 posts in 12 far right Facebook pages. It found "within online messaging, mobilisation frames were framed in specific ways to create certain narratives."[78] For example, the narrative that Muslims are 'inferior, sub-human, and inherently incompatible with Western liberal norms and values'; [79] or the narrative that 'gender fluidity and same sex marriage would open the gates for normalising paedophilia, polygamy, beastility [sic] and incest.'[80] These narratives are dangerous and need to be removed from digital platforms.

While there is a porous relationship between online hatred and extremist movements, policy parameters need to recognise that approaching this from a *countering violent extremism* angle alone will not address hate crime. A majority of the perpetrators of public acts of hatred both

---

72 Executive Council of Australian Jewry, above n 14, 10.

73 April Kailahi, Semisi Kailahi and Tatjana Bosevska, 'Resilient Women Project: Muslim Women and their experience of prejudice', (Project Report, Uniting Church of Australia Synod of Victoria and Tasmania, 4 June 2019).

74 Alan Berman and Shirleene Robinson, *Speaking Out: Stopping Homophobic and Transphobic Abuse in Queensland*, (Australian Academic Press, 2010).

75 Raphael Cohen-Almagor, 'Addressing Internet Dangerous Expressions: Deliberative Democracy and Cleanet ©' (2018) 21 *Journal of Internet Law* 11, 5-6.

76 Above n 30.

77 Puecker, Smith and Iqbal, above n 2.

78 Ibid 8.

79 Ibid 38.

80 Ibid 43.

offline and online, are ordinary people without affiliation to a far right group.[81] This means that the public advocacy of hatred or violence requires concerted policy treatment beyond only countering violent extremism, especially in the online sphere, to prevent and mitigate significant private and public harm.

81 Mason and Czapski, above n 17, 294.

Schedule 1

Relevant criminal offences that are almost never used in the online space

Urging violence against groups- *Criminal Code Act 1995 (Cth)* s80.2A

Publicly threatening or inciting violence- *Crimes Act 1900 (NSW)* s93Z

Serious vilification- *Criminal Code 2002 (ACT)* s750

Inciting racial animosity- *Criminal Code Amendment (Racial Vilification Act) 2004 (WA)* ss77-80

Serious racial and religious vilification- *Anti-Discrimination Act 1991 (QLD)* s131A; and

Serious vilification- *Racial and Religious Tolerance Act 2001* (VIC) ss24,25.

Schedule 2

Protected characteristics / categories of people protected

As of June 2018, the NSW Parliament amended the Crimes Act, deleting Section 20D (Offence of serious racial vilification), and adding in Section 93Z. This Section expanded the types and number of categories of people who were to be protected (protected characteristics) from only one category, race, to six categories of people who are targeted on the grounds of "race, religion, sexual orientation, gender identity or intersex or HIV/AIDS status".

Section 93Z provides the: "Offence of publicly threatening or inciting violence on grounds of race, religion, sexual orientation, gender identity or intersex or HIV/AIDS status"

93Z defines race, religion and sexual orientation in the legislation:

> "race includes colour, nationality, descent and ethnic, ethno-religious or national origin.
>
> religious belief or affiliation means holding or not holding a religious belief or view.
>
> sexual orientation means a person's sexual orientation towards:
> (a) persons of the same sex, or
> (b) persons of a different sex, or
> (c) persons of the same sex and persons of a different sex."

Section 93Z of NSW covers a broad range of those who are targeted with hate crimes.

In comparison, other jurisdictions in Australia and overseas also cover various protected characteristics.

Victoria Police:
"A prejudice motivated crime is a crime motivated by prejudice or hatred towards a person or a group because of a particular characteristic such as sexual orientation, gender identity, religion, race, sex, age, disability or homelessness. Many crimes can be motivated by prejudice, including harassment, threats, verbal abuse, destroying or damaging property, and in more serious cases, physical violence."
https://www.police.vic.gov.au/prejudice-and-racial-and-religious-vilification

UK Police:
"A hate crime is defined as 'Any criminal offence which is perceived by the victim or any other person, to be motivated by hostility or prejudice based on a person's race or perceived race; religion or perceived religion; sexual orientation or perceived sexual orientation; disability or perceived disability and any crime motivated by hostility or prejudice against a person who is transgender or perceived to be transgender.'"
https://www.met.police.uk/advice/advice-and-information/hco/hate-crime/what-is-hate-crime/

Canada:
Covers crimes motivated by hate, based on race, national or ethnic origin, language, colour, religion, sex, age, mental or physical disability, sexual orientation, or any other similar factor.
https://www150.statcan.gc.ca/n1/pub/85-002-x/2015001/article/14191-eng.pdf

USA FBI:
The FBI has defined a hate crime as a “criminal offense against a person or property motivated in whole or in part by an offender’s bias against a race, religion, disability, sexual orientation, ethnicity, gender, or gender identity.”
https://www.fbi.gov/investigate/civil-rights/hate-crimes

OSCE (Organization for Security and Cooperation in Europe):
The OSCE is composed of 37 countries. All countries have hate crime laws, but the protected characteristics vary; all hate crime laws in the OSCE include “race” as a protected category, (p37); the most commonly protected characteristics are “race”, national origin, and ethnicity; followed closely by religion (p40); almost all countries cover bias “motivated by religious or racial hatred”, while 11 countries have provisions extended to sexual orientation and 7 countries have provisions extended to disability, (p38); some countries include categories such as “gender,” “sexual orientation,” and “disability.” (p38).
https://www.osce.org/odihr/36426?download=true

Schedule 3

Table of exemptions to civil protections against racial hatred and vilification in Australia.

Act and Jurisdiction	Protected (Threshold-groups)	Exemptions
Racial Discrimination Act 1975 (Cth)	offend, insult, humiliate or intimidate - Race, national or ethnic origin, colour	*Exemptions s18D* Section 18C does not render unlawful anything said or done reasonably and in good faith: (a) in the performance, exhibition or distribution of an artistic work; or (b) in the course of any statement, publication, discussion or debate made or held for any genuine academic, artistic or scientific purpose or any other genuine purpose in the public interest; or (c) in making or publishing: (i) a fair and accurate report of any event or matter of public interest; or (ii) a fair comment on any event or matter of public interest if the comment is an expression of a genuine belief held by the person making the comment.
Racial and Religious Tolerance Act 2001 (VIC)	incites hatred against, serious contempt for, or revulsion or severe ridicule of- Racial and religious vilification	*Exceptions—public conduct (s11)* (1) A person does not contravene section 7 or 8 if the person establishes that the person's conduct was engaged in reasonably and in good faith— (a) in the performance, exhibition or distribution of an artistic work; or (b) in the course of any statement, publication, discussion or debate made or held, or any other conduct engaged in, for— (i) any genuine academic, artistic, religious or scientific purpose; or (ii) any purpose that is in the public interest; or (c) in making or publishing a fair and accurate report of any event or matter of public interest. *S. 11(2) inserted by No. 25/2006 s. 9.* (2) For the purpose of subsection (1)(b)(i), a religious purpose includes, but is not limited to, conveying or teaching a religion or proselytising.

Act and Jurisdiction	Protected (Threshold-groups)	Exemptions
Anti-Discrimination Act 1977 (NSW)	to incite hatred towards, serious contempt for, or severe ridicule of - Race, homosexual, HIV/AIDS status, transgender vilification	*S20C (Racial vilification exemptions)* (2) Nothing in this section renders unlawful-- (a) a fair report of a public act referred to in subsection (1), or (b) a communication or the distribution or dissemination of any matter on an occasion that would be subject to a defence of absolute privilege (whether under the *Defamation Act 2005* or otherwise) in proceedings for defamation, or (c) a public act, done reasonably and in good faith, for academic, artistic, scientific or research purposes or for other purposes in the public interest, including discussion or debate about and expositions of any act or matter. *s49ZT (homosexual vilification exemptions)* As above but includes 'religious instruction' as a purpose in (c) *49ZXB (HIV/AIDS vilification) and s38S (Transgender vilification) exemptions* As per racial vilification except 'religious discussion or instruction' is included as purpose in (c)
Discrimination Act 1991 (ACT)	incite hatred toward, revulsion of, serious contempt for, or severe ridicule of - Race, religious conviction, gender identity, sexuality, intersex status, disability, HIV/AIDS status vilification	s67A (2) However, it is not unlawful to— (a) make a fair report about an act mentioned in subsection (1); or (b) communicate, distribute or disseminate any matter consisting of a publication that is subject to a defence of absolute privilege in a proceeding for defamation; or (c) do an act mentioned in subsection (1) reasonably and honestly, for academic, artistic, scientific or research purposes or for other purposes in the public interest, including discussion or debate about and presentations of any matter.
Anti-Discrimination Act 1991 (QLD)	incite hatred towards, serious contempt for, or severe ridicule of - Race, religion, gender identity, sexuality vilification	s124A(2) *Subsection (1)* does not make unlawful— (a) the publication of a fair report of a public act mentioned in *subsection (1)* ; or (b) the publication of material in circumstances in which the publication would be subject to a defence of absolute privilege in proceedings for defamation; or (c) a public act, done reasonably and in good faith, for academic, artistic, scientific or research purposes or for other purposes in the public interest, including public discussion or debate about, and expositions of, any act or matter.

Act and Jurisdiction	Protected (Threshold-groups)	Exemptions
Civil Liability Act 1936 (SA)	inciting hatred, serious contempt or severe ridicule of a person or group of persons – race only	*S73 Racial victimisation* Act of racial victimisation… does not include (a) publication of a fair report of the act of another person; or (b) publication of material in circumstances in which the publication would be subject to a defence of absolute privilege in proceedings for defamation; or (c) a reasonable act, done in good faith, for academic, artistic, scientific or research purposes or for other purposes in the public interest (including reasonable public discussion, debate or expositions);
Anti-Discrimination Act 1988 (TAS)	offends, humiliates, intimidates, insults or ridicules - Age, Race, Disability, Sexual orientation, Lawful sexual activity, Gender, Gender identity, Intersex variations of sex characteristics, Pregnancy, Breastfeeding, Marital status, Relationship status, Family responsibilities, Parental status incite hatred towards, serious contempt for, or severe ridicule of - race, religion, sexual orientation, gender identity, intersex variations, disability	S55 The provisions of section 17(1) and section 19 do not apply if the person's conduct is – (a) a fair report of a public act; or (b) a communication or dissemination of a matter that is subject to a defence of absolute privilege in proceedings for defamation; or (c) a public act done in good faith for – (i) academic, artistic, scientific or research purposes; or (ii) any purpose in the public interest.